A Regularization Approach to the Minimum Distance Estimation : Application to Structural Macroeconomic Estimation Using IRFs *

Senay SOKULLU †

March 22, 2019

Abstract
This paper considers the invertibility problem of the optimal weighting matrix encountered during Impulse Response Function Matching Estimation (IRFME) of Dynamic Stochastic General Equilibrium (DSGE) Models. We propose to use a regularized inverse and derive the asymptotic properties of the estimator. We show that the asymptotic distribution of our estimator converges to that of the optimal estimator which has important implications for testing the fit of the model. We demonstrate the small sample properties of the estimator by Monte Carlo simulation exercises. Finally, we use our estimator to estimate the model in Altig et al.

JEL Classification: C10, E30

*I acknowledge helpful discussions with Jean-Pierre Florens, Engin Kara, Sami Stouli, Jon Temple and Frank Windmeijer. I also would like to thank Jean-Marc Robin and Richard Blundell, Francesco Zanetti and the two anonymous referees for their helpful comments.

†Department of Economics, University of Bristol, Priory Road Complex, Bristol, BS8 1TU, United Kingdom; e-mail: senay.sokullu@bristol.ac.uk

1 Introduction

The Minimum Distance Estimation (MDE) of structural VAR models, Impulse Response Function Matching Estimation (IRFME) has contributed a lot to the econometrics of DSGE models over the recent years. (See Rotemberg and Woodford, 1997; Giannoni and Woodford, 2004; Christiano et al., 2005; Altig et al., 2011). The method has gained favor as it is easy to implement and captures the dynamics of the model by construction. The parameters are estimated by minimization of the distance between the empirical IRFs obtained by a structural VAR and theoretical IRFs implied by the DSGE model. The optimal weighting matrix for the minimum distance estimation is given by the inverse of the variance-covariance matrix of the empirical impulse responses. However, estimation often suffers from the non-invertibility of this optimal weighting matrix because the inclusion of many horizons may lead to stochastic singularity. To solve this issue, instead of using the optimal weighting matrix, researchers either use an identity matrix like in Rotemberg and Woodford (1997) or a diagonal matrix with the inverse variances of the IRFs on the diagonal as in Christiano et al. (2005) or Altig et al. (2011). In this paper, we address this problem and propose the use of a regularized inverse of the covariance matrix of the IRFs. To the best of our knowledge, this is the first time that regularization is proposed to solve the problem of invertibility of the optimal weighting matrix in IRFME.[1]

We know from econometric theory that the minimum distance estimator gives consistent estimates for any weighting matrix that is positive definite and symmetric, see Gourieroux et al. (1995). So the use of a non optimal weighting matrix does not affect the consistency of the estimated parameters of the structural macroeconomic model. However, it results in less efficient estimates, i.e., larger variances. Moreover, as pointed out by Feve et al. (2009), the use of a non optimal weighting matrix may lead to erroneous conclusions while testing the fit of the model as the J-statistic is not asymptotically distributed as a χ^2 with degrees of freedom given by the model. Feve et al. (2009) propose to use a simulated distribution to test the fit of the model. We

[1] Although our focus is on IRFME of DSGE models in this paper, there are other econometric methods which are used in the empirical analysis of DSGE models. For example, maximum likelihood methods have been used to estimate DSGE models; the examples may go back as far as Christiano (1988) and Altug (1989). More recent examples include Ireland (2004) and Zanetti (2008), among others. Another popular method for the estimation of DSGE models is the Bayesian analysis, see An and Schorfheide (2007); Mandelman and Zanetti (2014). For more details on different estimation methods of DSGE models, see Fernández-Villaverde et al. (2016).

show that our proposed estimator has the same asymptotic distribution as the optimal one under some regularity conditions. Hence there is no need to use a simulated distribution to test the model's fit.

The number of horizons to be included in the DSGE model is also closely related to the efficiency of the estimator. Hall et al. (2012) propose two new criteria to be used in IRFME. The first one is the Valid Impulse Response Selection Criterion (VIRSC) and the second one is the Relevant Impulse Response Selection Criterion (RIRSC). The former gives information on which impulse response functions to include in order to obtain consistent estimators whilst the latter excludes redundant impulse responses and provides the set of valid ones. In this paper, we are not concerned with the optimal number of horizons and our estimator, by construction, can deal with an infinite number of horizons. The simulation results show that with the higher number of horizons, the performance of our estimator relative to that of the optimal estimator with generalized inverse improves.

This problem with IRFME is very similar to the problem of many instruments in the econometrics literature. Although increasing the number of instruments improves the asymptotic efficiency of the estimator as is shown by Carrasco and Florens (2014), in finite samples, increasing the number of instruments increases the bias, see Bekker (1994) and Hansen et al. (2008), and it may lead to a poor Gaussian asymptotic approximation, see Newey and Windmeijer (2009). Moreover, as in the case of IRFME it may lead to a non-invertible covariance matrix of the moment conditions. The problem can be dealt with by including fewer instruments but this will result in an efficiency loss. One solution is proposed by Donald and Newey (2001). They propose a selection rule for the instruments based on minimization of the approximate MSE. However, their method does not work well in 2SLS estimation when we do not have a priori information on the relative importance of instruments. Other proposed solutions use empirical likelihood techniques, see Donald et al. (2003) and Kitamura et al. (2004). Carrasco et al. (2007b) specify the estimation with many instruments as an inverse problem and point out the ill-posedness which is the result of a non-invertible covariance matrix. Following this idea, they propose to use the regularized inverse of the covariance matrix. Moreover, Carrasco and Florens (2000) and Carrasco (2012) show that the resulting regularized IV estimator is consistent and asymptotically normal. In this paper, we follow Carrasco and Florens (2000) and Carrasco (2012) and treat the problem as an ill-posed inverse problem and propose to regularize the covariance matrix of the IRFs by *Tikhonov Regularization*. We show that the regularized IRFME is consistent and asymptotically normally distributed. Additionally, we simulate a simple Real Business Cycle (RBC) model and show that with the optimal choice of the regularization parameter, we obtain better results in terms of bias and standard deviation, than the cases where the identity or diagonal weighting matrices are used.

One issue with using a regularized inverse is the selection of the regularization parameter which acts like a smoothing parameter. Although the choice is quite straightforward with simulated data, as we can pick the one that minimizes the MSE, it may not be that easy when we work with real data. We estimate the model in Altig et al. (2011) by using a regularized inverse of the optimal weighting matrix in which we use a data driven selection rule for the regularization parameter. From a grid of values we pick the one that minimizes the sum of the norm of the estimated standard errors of the parameters and the norm of the distance between the empirical and theoretical IRFs. Our estimation results are in line with econometric theory in the sense that the estimated parameters are close to the ones obtained in Altig et al. (2011) but the variances of the parameters are smaller.

The paper proceeds as follows. In Section 2 we define IRF Matching Estimators. In Section 3, we introduce the use of Tikhonov Regularization in IRFME while in Section 4 we show its asymptotic properties. A Monte Carlo simulation and its results are presented in Section 5, while we present our application example in Section 6. Finally in Section 7, we conclude. All of the proofs are presented in the Online Appendix.

2 IRF Matching Estimator

In this section we define the IRF matching estimator that is used to estimate structural parameters in DSGE models. Note that the IRF matching estimator is a form of Indirect Inference estimators developed by Gourieroux et al. (1993). Later Dridi et al. (2007) considered the misspecification in DSGE models estimated by Indirect Inference techniques, but throughout this paper we assume that the model is correctly specified.

Let X_t be the $(d_X \times 1)$ vector of variables of interest at date $t = 1, 2, ..., T$. The IRF matching estimator is based on the minimum distance estimator which minimizes the distance between the empirical IRFs obtained by fitting a VAR model to X_t and the theoretical IRFs implied by the structural model. Suppose that the VAR model is:

$$X_t = \Gamma_0 + \Gamma_1 X_{t-1} + ... + \Gamma_s X_{t-s} + \epsilon_t \tag{1}$$

where $s \geq 1$ and $\epsilon_t \sim iid(0, \Omega)$. We want the VAR model given in (1) to have an infinite order VMA representation and IRFs. Hence, we assume that

$$\Gamma(L) = I_{d_X} - \Gamma_1 L - \Gamma_2 L^2 - ... - \Gamma_s L^s$$

is invertible where L is the lag operator, I_{d_X} is a $(d_X \times d_X)$ identity matrix and s is finite.

Let $\hat{\varphi}_h$ denote the vector of estimated IRFs from the VAR model up to horizon h and $\psi_h(\theta)$ denote the IRFs obtained from the structural model up to horizon h.[2] Then the IRF matching estimator is given by:

$$\hat{\theta}_h = \underset{\theta \in \Theta}{\operatorname{argmin}} [\hat{\varphi}_h - \psi_h(\theta)]' A [\hat{\varphi}_h - \psi_h(\theta)] \tag{2}$$

where A is a $q \times q$ symmetric and positive semi definite weighting matrix where $q = d_X^2 \times h$. Under standard regularity conditions, one can show that the estimator $\hat{\theta}_h$ is consistent and asymptotically normal, i.e:

$$\hat{\theta}_h \xrightarrow{p} \theta_0$$

and

$$\sqrt{T}(\hat{\theta}_h - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma_0)$$

where

$$\Sigma_0 = \left(\frac{\partial \psi_h'}{\partial \theta}(\theta_0) A \frac{\partial \psi_h}{\partial \theta'}(\theta_0) \right)^{-1} \left(\frac{\partial \psi_h'}{\partial \theta}(\theta_0) A W A \frac{\partial \psi_h}{\partial \theta'}(\theta_0) \right) \left(\frac{\partial \psi_h'}{\partial \theta}(\theta_0) A \frac{\partial \psi_h}{\partial \theta'}(\theta_0) \right)^{-1} \tag{3}$$

where W is the covariance matrix of the IRFs. If we choose the weighting matrix A to be equal to the inverse of the covariance matrix of the IRFs, we get the optimal estimator, i.e., we get the smallest variance for $\hat{\theta}$, see Gourieroux et al. (1995); Ruud

[2]For computation of IRFs, see Hamilton (1994), Chapter 11.

(2000). Then its asymptotic variance is given by:

$$\Sigma_0 = \left(\frac{\partial \psi_h'}{\partial \theta}(\theta_0) W^{-1} \frac{\partial \psi_h}{\partial \theta'}(\theta_0) \right)^{-1}$$

As pointed out in Feve et al. (2009), in many applications the number of IRFs included is much larger than the number of estimated parameters in the structural VAR and this may lead to the problem of invertibility for the optimal weighting matrix. In the next section we present our method to deal with this problem of invertibility.

3 Tikhonov Regularization for IRFME

Let us re-write the estimation problem such that it later becomes easier to state the assumptions to get the asymptotic properties of the estimator. Let $f(h, \theta) \in \mathbb{R}^q$ define the distance between the empirical IRFs and the IRFs implied by the structural DSGE model up to horizon h:

$$f(h, \theta_0) = \varphi_h - \psi_h(\theta_0) \tag{4}$$

where φ_h is the vector IRFs coming from the structural VAR and depends on the process $\{X_t\}$, $X_t \in \mathbb{R}^{d_X}$ and $\psi_h(\theta_0)$ is the vector of IRFs implied by the structural DSGE model where θ_0 is the true value of the vector of parameters. Note that q depends on h as well as the dimension of the process $\{X_t\}$, $q = h \times d_X^2$. Moreover, q might be smaller or larger than the sample size. In this section, the only assumption we make related to the dimension of q is that h is finite.[3]

Let $\mathcal{E}$ be the Hilbert space corresponding to $\mathbb{R}^q$ and let A be a linear operator defined on $\mathcal{E}$. Note that since $\mathcal{E}$ is finite dimensional, A is a matrix and the adjoint operator of A, A^* is equal to its transpose A', i.e., $A^* = A'$, see Carrasco et al. (2007b).

The estimator of θ_0 is given by:

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \hat{f}'(h, \theta) A_T' A_T \hat{f}(h, \theta) \tag{5}$$

where A_T is a sequence of linear operators converging to A and $\hat{f}(.)$ is the empirical

[3]The case h infinite will be discussed in Appendix B.

counterpart of Equation (4):

$$\hat{f}(h,\theta) = \hat{\varphi}_h - \psi_h(\theta)$$

The estimator obtained in Equation (5) is consistent for any A_T converging to A and it is efficient if $A_T' A_T$ is a consistent estimator of the inverse of the covariance matrix of IRFs, W^{-1}. In cases where the number of IRFs exceeds the number of estimated parameters in the structural VAR, we have a problem like that of many instruments in IV regression.[4] In this case the estimator of W, $\hat{W}$ becomes near singular or singular and thus non-invertible. Thus the estimation becomes an ill-posed inverse problem, i.e. the solution is not continuous.[5] In particular, with IRFME problems, when the dimension of ϵ_t is smaller than d_X, W itself becomes singular, see footnote 3 of Hall et al. (2012). Hence, in this paper we do not rule out singular W and we use generalized inverse of W in our proofs. In the case of IV with many instruments where the variance-covariance matrix is nearly singular, Carrasco and Florens (2000), Carrasco et al. (2007b) and Carrasco (2012) propose to solve this ill-posed inverse problem by regularization methods. Following this literature, instead of using non-optimal weighting matrices, we propose to regularize the inverse using the *Tikhonov Regularization* scheme. Under this scheme, the regularized inverse of the weighting matrix is given by:[6]

$$\hat{W}^{\alpha} = (\alpha I_q + \hat{W}'\hat{W})^{-1}\hat{W} \tag{6}$$

where $\alpha > 0$ and $\alpha \xrightarrow{T\to\infty} 0$ is called the *regularization parameter* and I_q is a $(q \times q)$ identity matrix.

It is clear from equation (6) that, on the one hand, we are perturbing the optimal weighting matrix by adding α to its diagonal, i.e., adding a source of bias to the

[4] See Carrasco and Florens (2000), Donald and Newey (2001), Donald et al. (2003), Hansen et al. (2008), Newey and Windmeijer (2009).

[5] As defined in Engl et al. (1996), a problem is well-posed if the definitions below hold:

(i) For all admissible data a solution exist.

(ii) For all admissible data the solution is unique.

(iii) The solution continuously depends on the data.

[6] Note that W becomes an operator when we use $h = \infty$ and the regularized inverse is given by $(W^{\alpha})^{-1} = (\alpha I + W^*W)^{-1}W^*$

estimation. On the other hand, regularization of the inverse of the covariance matrix leads to a more stable solution and thus decreases the variance of the weighting matrix. So α acts like a smoothing parameter and its choice is very important as it balances the trade-off between the fit and the variance. The main rule adopted for the selection of the regularization parameter in nonparametric problems is called 'discrepancy principle' and it is based on minimizing the noise, see Florens and Sokullu (2017). In parametric problems, the optimal regularization parameter is given by the minimizer of the MSE. However, since the regularized estimator is consistent and has the same asymptotic distribution as the optimal estimator, we need to do a second order analysis.[7] The choice of α is studied by Carrasco (2012) in the case of the many instruments problem. After obtaining an approximation for the second order expansion of the MSE, she derives a formula for the optimal α.[8] However, as pointed out in Carrasco et al. (2007a), this becomes very complicated in a time series context. For that reason we propose a data-based selection criteria which depends on the fit of the model as well as the standard errors of the estimated parameters. Hence, the data driven optimal α is given by:

$$\alpha_{opt} = \underset{\alpha \in \mathcal{G}}{\operatorname{argmin}} \|\hat{\varphi}_h - \psi_h(\hat{\theta}^\alpha)\|^2 + \|\hat{\sigma}(\hat{\theta}^\alpha)\|^2 \tag{7}$$

where $\mathcal{G}$ is a grid of α, $\hat{\theta}^\alpha$ is the vector of estimated parameters with Tikhonov regularized estimation and $\hat{\sigma}(\hat{\theta}^\alpha)$ is the vector of standard error of the estimated parameters. This criterion captures two important aspects: First, it can be considered as a function of the noise which is in line with the nonparametric approach. Secondly, as we propose to minimize the sum of the Euclidian norm of the fit and the standard errors of the estimated parameters, it can also be seen as an approximation to the MSE.

It should be noted that there are also other regularization schemes in inverse problems literature, but in this paper we adopt Tikhonov Regularization. Carrasco (2012) considers 3 regularization schemes: Tikhonov, Landweber-Friedman and Spec-

[7] In the next section we show that it has the same asymptotic distribution as the optimal one. Moreover, it is also proved by Carrasco and Florens (2000) and Carrasco (2012) for regularized GMM estimation with many moment conditions.

[8] Though the choice of optimal α may not be the same for all inverse problems. The criterion function to be minimized by the selection of α changes from application to application. See Golub et al. (1979) and Engl et al. (1996).

tral Cut-off (she also considers Principal Components which is a variation of Spectral Cut-off). There are several reasons that we focus on Tikhonov regularization in this paper. First of all, it is closely related to Ridge Regression so it is easier to be understood by a general audience. Secondly, for linear problems it is more practical to use Tikhonov regularization rather than using iterative Landweber-Friedman regularization and there is little difference in terms of finite sample properties of the estimators obtained either with Tikhonov or Landweber-Friedman regularization schemes, see Carrasco (2012). Finally, Monte Carlo experiments in Carrasco (2012) show that Principal Component performs always worse than Tikhonov regularization.

4 Asymptotic Properties of the Regularized IRFME

In this section we show that the regularized IRFME is consistent and asymptotically normal. Moreover, we show that its asymptotic distribution is equal to that of the optimal estimator. The consistency of the regularized estimator in case of many instruments in GMM has been shown before by Carrasco and Florens (2000) and Carrasco (2012), nonetheless here we prove it for the impulse response function matching estimation.[9]

We mentioned in the previous sections that our solution to the non-invertibility problem in IRFME holds even if the horizon is chosen to be infinite, i.e. $h = \infty$. If $h = \infty$ then we work with infinite dimensional operators. However, in DSGE models the horizon is taken to be finite. Hence, we state our assumptions and the asymptotic properties of the regularized IRFME for $h < \infty$. The results for $h = \infty$ are presented in Appendix B.

The following assumptions are needed for the consistency and asymptotic normality:

Assumption 1 *Let X be a stationary, ergodic, random process. $\{X_t\}_{t=1}^{T}$, $X_t \in \mathbb{R}^{d_X}$ is an observed random sample of X.*

Assumption 2 *(i) Let φ_h be the IRFs up to horizon h from the structural VAR. $\varphi_h : \mathbb{R}^r \to \mathcal{E}$, where r is the number of existing moments of $\{X_t\}_{t=1}^{T}$ and $\mathcal{E}$ is a Hilbert space corresponding to $\mathbb{R}^q$ as defined before and it is endowed with the Euclidian norm.*

[9] The results of Carrasco and Florens (2000) hold for the case of a continuum of moment conditions.

Moreover, assume that a consistent estimator of φ_h, $\hat{\varphi}_h$ exists and $\sqrt{T}(\hat{\varphi}_h - \varphi_h) \xrightarrow{d} \mathcal{N}(0, W)$
(ii) Let $\psi_h(\theta)$ be a function of theoretical IRFs from the structural DSGE model up to horizon h. $\psi_h : \mathbb{R}^k \to \mathcal{E}$, where $\theta \in \Theta \in \mathbb{R}^k$ is the vector of parameters to be estimated.

Assumption 3 *The function $f : \mathcal{E} \to \mathcal{E}$ is defined as:*

$$f(h, \theta) = \varphi_h - \psi_h(\theta)$$

and it is equal to 0 for $\theta = \theta_0$.

Assumption 4 *Let A be a nonrandom bounded linear operator $A : \mathcal{D}(A) \subset \mathcal{E} \mapsto \mathcal{E}$ and let $A'A = W^+$ where W^+ is the generalized inverse of the covariance matrix of φ_h, W. We assume that $f(h, \theta) \in \mathcal{D}(W^+)$ for all θ.*

Assumption 5 *Let $N(A)$ denote the null space of A: $N(A) = \{g \in \mathcal{E} | Ag = 0\}$. We assume that $f(h, \theta) \in N(A)$ implies $f(h, \theta) = 0$.*

Assumption 6 *$f(h, \theta)$ is differentiable with respect to θ and $\frac{\partial f(h,\theta)}{\partial \theta'} \in \mathcal{D}(W^+)$ is full rank.*

Assumption 7 *Let A_T be a sequence of random bounded linear operators. $A_T : \mathcal{D}(A_T) = \mathcal{E} \mapsto \mathcal{E}$. Let $\hat{f}(h, \theta) = \hat{\varphi} - \psi(\theta)$. We assume that $\hat{f}(h, \theta) \in \mathcal{D}(A_T) \quad \forall \theta$ and $Q_T = \left\| A_T \hat{f}(h, \theta) \right\|$ is a continuous function of θ.*

Assumption 8 *$Q_T \to Q = \|Af(h, \theta)\|$ almost surely on $\theta \in \mathbb{R}^k$*

Assumption 9 *The $(k \times k)$ matrix $\frac{\partial \hat{f}'(h,\hat{\theta})}{\partial \theta} A'A \frac{\partial \hat{f}(h,\hat{\theta})}{\partial \theta'}$ is positive definite and symmetric.*

Assumption 10 *A and A_T commute with differential operator:*

$$\frac{\partial}{\partial \theta'}[Au(\theta)] = A[\frac{\partial}{\partial \theta'} u(\theta)]$$

where $u(\theta)$ is any function of θ.

Assumption 11 $\mathbb{E} \left\| f(h, \theta_0) \right\|^4 < \infty$

Assumption 12

$$\left\|\hat{f}(h,\theta)-f(h,\theta_0)\right\| = O_p\left(\frac{1}{\sqrt{T}}\right)$$

$$\left\|\frac{\partial \hat{f}'}{\partial\theta}(h,\theta)-\frac{\partial f'}{\partial\theta}(h,\theta_0)\right\| = O_p\left(\frac{1}{\sqrt{T}}\right)$$

Assumptions 1 to 4 define the stochastic process X and the estimation problem. Assumption 5 is a condition for identification. In its simplest form, Assumption 5 can be a full rank condition on $A'A$. However, note that this assumption does not rule out the case that $N(A)$ has more elements than 0, which indeed allows for the cases where $A'A$ is singular. Assumption 6 is also needed for identification. It indeed guarantees the first order identification which may fail in nonlinear problems, see Sargan (1983). As was pointed out by Dovonon and Renault (2013) and Dovonon and Gonçalves (2014), the lack of first order identification has important implications for the test of overidentifying restrictions. Assumption 7 guarantees the continuity of the objective function in θ whereas Assumption 8 implies that the empirical weighted distance converges to its population value. Assumptions 9 to 12 are further conditions needed to prove the consistency and asymptotic distribution of the regularized estimator we propose. Note that the assumptions we have stated are mostly standard in minimum distance estimation. There is only one modification which appears in Assumption 5 as we allow for singularity for the covariance matrix of IRFs.

Lemma 1 *Let $W^\alpha = (\alpha I + W'W)^{-1}W'$ be the Tikhonov regularized inverse of the operator W such that $A'A = W^{-1}$ if W is nonsingular and $A'A = W^+$ otherwise where W^+ is the generalized inverse of W. Assume that Assumption 4 holds. Moreover let $\hat{W}^\alpha = (\alpha I + \hat{W}'\hat{W})^{-1}\hat{W}'$ be the regularized inverse $\hat{W}$, $A_T'A_T = \hat{W}^{-1}$ (or $A_T'A_T = \hat{W}^+$ if $\hat{W}$ is singular) and $\hat{W}$ converges to W, Then:*

$$\left\|\hat{W}^\alpha \hat{f}(h,\theta) - W^+ f(h,\theta_0)\right\| \to 0 \quad \textit{in probability as} \quad T\to\infty, \alpha^3 T\to\infty, \alpha\to 0$$

In IRFME the optimal weighting matrix might indeed be singular. In such a case we may need to use generalized inverse as the theory about optimality follows with the generalized inverse as well. Hence in Lemma 1, we show that the empirical weighted distance we want to minimize converges to the population distance where the weighting matrix is the generalized inverse of the covariance matrix of the IRFs.

Theorem 2 *Under the Assumptions (1) to (11), the estimator:*

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \hat{f}'(h,\theta)\hat{W}^{\alpha}\hat{f}(h,\theta)$$

satisfies:
(i) $\hat{\theta} \to \theta_0$ *in probability as* $T \to \infty$, $\alpha^3 T \to \infty$ *and* $\alpha \to 0$
(ii) $\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, \left(\frac{\partial f}{\partial \theta}(h,\theta_0) W^{+} \frac{\partial f}{\partial \theta'}(h,\theta_0)\right)^{-1}\right)$ *as* $T \to \infty$, $\alpha^3 T \to \infty$ *and* $\alpha \to 0$

Theorem (2) states that the regularized IRFME estimator is consistent and optimal.

Remark 3 *The problem of invertibility of the weighting matrix is not treated by the use of a generalized inverse in any of the aforementioned DSGE literature. Although, it would lead to the same asymptotic distribution as the regularized IRFME, we show in Section 5, by means of a Monte Carlo simulation that the estimator obtained with the generalized inverse is not as stable as the regularized estimator.*

5 Simulation

In this section we conduct a Monte Carlo simulation to compare the small sample performance of the estimator we propose with the small sample performance of other estimators used in the IRFME literature. We compare the estimators which are obtained from four different approaches. In the first approach we use the one that we propose, i.e., the covariance matrix of the IRFs is inverted by regularizing it with a Tikhonov Regularization scheme. The second approach uses a diagonal weighting matrix which has the inverse of the variances of the IRFs on the diagonal whereas the third one uses an identity matrix as the weighting matrix. Finally the fourth method uses the generalized inverse of the covariance matrix of the IRFs as the weighting matrix.

5.1 The Simulation Design

We adopt the simulation design given in Section 7.2 in Hall et al. (2012).[10] It is a simple RBC model which can be represented by a bivariate VAR(3) in labour productivity (y_t/l_t) and employment (l_t). The model follows from Watson (2006). Watson (2006) derives the VAR representation of a simplified version of the RBC model presented in Christiano et al. (2006) where labour productivity can only be affected by a technology shock in the long run. Moreover, Christiano et al. (2006) consider two versions of their model. The first version - *standard* or *non-recursive version* - makes the standard timing assumption: Time t decisions are taken after time t shocks are realized. The second version - *recursive version* - assumes that in the first stage agents make their labour supply decisions after observing the tax on labour income. In the second stage, all other shocks are realized and investment and consumption decisions are made.[11]

Watson (2006) simplifies the model in Christiano et al. (2006) by suppressing the constant and the terms involving capital stock. He then derives a model which uses the impulse responses of labour productivity and employment to shocks to technology and labour income tax. Following Watson (2006), the model we use in our simulation design is given by:

$$
\begin{aligned}
\begin{pmatrix} 1 & \alpha_y \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta ln\left(\frac{y_t}{l_t}\right) \\ ln(l_t) \end{pmatrix} &= \begin{pmatrix} 0 & \alpha_y \\ \gamma_1 & \rho_l - \alpha_y\gamma_1 \end{pmatrix} \begin{pmatrix} \Delta ln\left(\frac{y_{t-1}}{l_{t-1}}\right) \\ ln(l_{t-1}) \end{pmatrix} \\
&+ \begin{pmatrix} 0 & 0 \\ -\rho_l\gamma_1 & \alpha_y(1+\rho_l)\gamma_1 \end{pmatrix} \begin{pmatrix} \Delta ln\left(\frac{y_{t-2}}{l_{t-2}}\right) \\ ln(l_{t-2}) \end{pmatrix} \\
&+ \begin{pmatrix} 0 & 0 \\ 0 & \alpha_y\rho_l\gamma_1 \end{pmatrix} \begin{pmatrix} \Delta ln\left(\frac{y_{t-3}}{l_{t-3}}\right) \\ ln(l_{t-3}) \end{pmatrix} + \begin{pmatrix} \eta_t \\ \nu_t \end{pmatrix}
\end{aligned}
$$

where α_y is share of capital in production, ρ_l is the serial correlation coefficient in the labour income tax process. Moreover let $\tilde{a}_z$ and a_z be the parameters associated with the lagged state of technology in the policy rule for labour in the standard

[10] We adopt exactly the same design as in Hall et al. (2012) so that we can have a direct comparison of our results.

[11] We briefly present the model in Christiano et al. (2006) in Appendix C. For details of the model, see Christiano et al. (2006).

and recursive models, respectively.[12] Then one can write $\gamma_1 = (\tilde{a}_z - a_z\rho_l)/(1-\alpha_y)$; $\eta_t = (1-\alpha_y)\sigma_z\epsilon_t^z$; $\nu_t = \alpha_l\sigma_l\epsilon_t^l$ and ϵ_t^z and ϵ_t^l are i.i.d. zero mean unit variance shocks.

The structural parameters to be estimated are α_y, ρ_l and γ_1. Hall et al. (2012) assigned the following values to the parameters of the model: $\alpha_y = 0.35, \sigma_z = 1, \sigma_l = 1, \rho_l = 0.95, \alpha_l = 1, a_z = 0, \tilde{a}_z = 0.325$ and $\gamma_1 = 0.5$. Following Hall et al. (2012), we simulate the model 1000 times for sample sizes 100, 200 and 400. The horizon h is set to 12.[13] For the Tikhonov regularized estimator, we use the data-based selection rule as defined in Section 3 to choose the regularization parameter, α. We set a grid of values for α and estimate the regularized model for each value on the grid. We select α as the one that minimizes the sum of the Euclidian norm of the fit and the Euclidian norm of the standard deviations of the estimated parameters where the standard deviations are calculated using optimal variance formula. Since in the application of DSGE models the econometrician is interested in getting the best fit with the best precision, we find this selection rule intuitive. The grid we use for α in the simulations is an equidistant log-space of 40 points in the interval $[-9, 0]$. The simulation is performed by user written code in MATLAB where we modified the base code used in Hall et al. (2012).[14]

5.2 Results

The results are presented in Tables 1 to 6. Following Hall et al. (2012), we report the median absolute bias, standard deviation and coverage probabilities of the 95% confidence interval based on the t-test, holding the number of impulse responses fixed.[15] Moreover, in all tables we present results for different number of horizons, $h \in [1, 12]$ where h indicates that impulse responses up to horizon h - including h - are used in the estimation. Tables 1 - 3 show the results obtained with the

[12] See equation (2) in Watson (2006).

[13] Note that as mentioned in Hall et al. (2012), in this simulation design the problem becomes ill-posed once we set $h > 3$.

[14] We thank Barbara Rossi for making the codes available on her website.

[15] For an estimate $\hat{\beta}$ of the true value β_0:

- Median absolute bias: $Med[|\hat{\beta} - \beta_0|]$
- Standard deviation: $E[(\hat{\beta} - \beta_0)^2]^{1/2}$
- Coverage probability: Proportion of the time that the estimated confidence interval contains β_0

regularized weighting matrix. The first thing to note is that the results improve with the sample size, i.e. bias and standard deviation become smaller and coverage probability increases. Moreover since the ill-posedness of the problem increases with the number of impulse responses for a given sample size one would expect that the bias and the standard deviation increase and the coverage becomes less accurate as the number of horizons increases. However, thanks to regularization, bias does not increase drastically and the standard deviation decreases with the number of horizons although coverage probability does worsen slightly. Especially, compared to Tables 4a to 4c in Hall et al. (2012), the regularized estimator performs very well.

Table 1: **Simulation Results with Regularized Weighting Matrix, T=100**

Bias, Standard Deviation, Coverage Probability									
	α_y			γ_1			ρ_l		
h	bias	std dev	prob	bias	std dev	prob	bias	std dev	prob
1	0.004	0.088	0.951	0.000	0.108	0.942	-0.005	0.138	0.933
2	-0.005	0.068	0.941	0.001	0.063	0.945	-0.013	0.075	0.932
3	-0.005	0.065	0.939	-0.003	0.073	0.955	-0.016	0.058	0.931
4	-0.006	0.065	0.932	-0.002	0.066	0.944	-0.016	0.051	0.934
5	-0.008	0.066	0.919	-0.004	0.073	0.941	-0.015	0.046	0.938
6	-0.007	0.067	0.911	-0.003	0.071	0.928	-0.015	0.044	0.919
7	-0.006	0.069	0.899	-0.003	0.074	0.927	-0.014	0.042	0.904
8	-0.006	0.070	0.897	-0.004	0.072	0.922	-0.014	0.042	0.888
9	-0.005	0.070	0.884	-0.003	0.074	0.920	-0.013	0.041	0.875
10	-0.005	0.070	0.888	-0.003	0.073	0.919	-0.013	0.041	0.865
11	-0.006	0.070	0.889	-0.004	0.074	0.921	-0.013	0.041	0.857
12	-0.008	0.069	0.892	-0.003	0.073	0.921	-0.013	0.041	0.850

Simulation results for different number of horizons, h. Estimated parameters: Share of capital in production, α_y, AR(1) coefficient of the labour income tax process, ρ_l and the effect of change in labour productivity on employment, γ_1.
Source: Author's calculations

Table 4 shows the results with a diagonal weighting matrix while the results obtained with the identity weighting matrix are given in Table 5. Finally, in Table 6 we present the results obtained with the generalized inverse of the covariance matrix of the IRFs.[16] The results obtained with generalized inverse are significantly worse than the other results, especially in terms of coverage probability. Although theoretically

[16]For the sake of exposition, in this section we report the results obtained with diagonal, identity and generalized inverse weighting matrices only for the sample size T=200. Results for other sample sizes are available upon request from the author.

Table 2: **Simulation Results with Regularized Weighting Matrix, T=200**

Bias, Standard Deviation, Coverage Probability									
	α_y			γ_1			ρ_l		
h	bias	std dev	prob	bias	std dev	prob	bias	std dev	prob
1	0.004	0.060	0.953	-0.001	0.072	0.948	-0.005	0.092	0.955
2	-0.002	0.045	0.953	-0.001	0.042	0.962	-0.002	0.049	0.949
3	-0.003	0.042	0.958	-0.002	0.049	0.974	-0.004	0.038	0.954
4	-0.003	0.042	0.955	-0.002	0.045	0.973	-0.006	0.034	0.949
5	-0.003	0.044	0.955	-0.003	0.049	0.969	-0.007	0.030	0.952
6	-0.002	0.044	0.952	-0.004	0.049	0.963	-0.007	0.028	0.945
7	-0.003	0.045	0.949	-0.003	0.051	0.962	-0.007	0.027	0.943
8	-0.003	0.045	0.945	-0.004	0.049	0.960	-0.007	0.026	0.930
9	-0.003	0.045	0.937	-0.003	0.049	0.955	-0.006	0.026	0.928
10	-0.003	0.044	0.936	-0.003	0.048	0.956	-0.006	0.026	0.923
11	-0.003	0.044	0.933	-0.003	0.048	0.952	-0.007	0.026	0.908
12	-0.003	0.044	0.930	-0.004	0.048	0.956	-0.007	0.026	0.895

Simulation results for different number of horizons, h. Estimated parameters: Share of capital in production, α_y, AR(1) coefficient of the labour income tax process, ρ_l and the effect of change in labour productivity on employment, γ_1.
Source: Author's calculations

one can consider the generalized inverse of the covariance matrix of the IRFs as the optimal one, in practice it leads to unstable results. Indeed this may explain why in practice this optimal weighting matrix is not used. Moreover, standard deviation of the estimates obtained with the regularized weighting matrix are smaller than those of the estimates obtained with identity or diagonal weighting matrices though bias of the estimate of ρ_l is higher. The coverage probabilities under the latter two methods are larger than the coverage probabilities under the former; although the magnitude is very small. The biggest difference in coverage probability is for the estimate of the ρ_l. Since the identity and diagonal weighting matrices do not suffer from the stochastic singularity, increasing the number of horizons does not have the detrimental effect as it has on the estimates obtained with generalized inverse of the covariance matrix of IRFs. Although, as the simulation results show, regularization introduces stability to the inverse and improves the results. Moreover, as already mentioned, the distribution of the regularized estimator converges to the that of the optimal one, hence the J-statistic to test the fit of the model converges the χ^2 (with the degrees of freedom given by the model). Feve et al. (2009) point out that J-statistic obtained by the use of diagonal or identity weighting matrices may not necessarily converge

Table 3: **Simulation Results with Regularized Weighting Matrix, T=400**

Bias, Standard Deviation, Coverage Probability									
	α_y			γ_1			ρ_l		
h	bias	std dev	prob	bias	std dev	prob	bias	std dev	prob
1	-0.000	0.042	0.944	-0.004	0.051	0.962	-0.000	0.067	0.942
2	0.001	0.032	0.947	-0.001	0.029	0.973	-0.003	0.036	0.943
3	-0.000	0.031	0.956	0.000	0.034	0.979	-0.003	0.028	0.943
4	0.000	0.030	0.959	-0.001	0.031	0.978	-0.004	0.025	0.936
5	-0.001	0.031	0.965	-0.000	0.035	0.985	-0.003	0.022	0.936
6	-0.001	0.032	0.963	-0.001	0.034	0.982	-0.004	0.020	0.932
7	-0.001	0.032	0.962	-0.001	0.036	0.980	-0.003	0.019	0.935
8	-0.001	0.032	0.955	-0.002	0.035	0.975	-0.003	0.019	0.930
9	-0.001	0.032	0.952	-0.001	0.035	0.970	-0.003	0.018	0.925
10	0.000	0.032	0.948	-0.001	0.035	0.965	-0.003	0.018	0.920
11	0.000	0.032	0.945	-0.002	0.034	0.967	-0.003	0.018	0.907
12	-0.000	0.031	0.942	-0.001	0.034	0.969	-0.003	0.018	0.905

Simulation results for different number of horizons, h. Estimated parameters: Share of capital in production, α_y, AR(1) coefficient of the labour income tax process, ρ_l and the effect of change in labour productivity on employment, γ_1.
Source: Author's calculations

to χ^2 and they propose to use a simulated distribution. Thus, one can say that the use of regularized weighting matrix brings about a trade-off: slightly poorer coverage probability vs. optimal asymptotic distribution which makes the inference easier as well as more reliable compared to the estimators using identity or diagonal matrices.

In their simulation exercise Hall et al. (2012) select the relevant and valid impulse responses using the generalized inverse of the covariance matrix of the IRFs as the weighting matrix. Although the coverage probabilities increase compared to the case where all impulse responses are used, they are still smaller then the coverage probabilities obtained with the regularized inverse.[17]

Finally, note that all the results we present here hold for the regularized estimator with the optimal regularization parameter α and it is selected as the one that minimizes the sum of the Euclidian norm of fit and standard deviations of the estimated parameters. In practice, α depends on the sample size and on the degree of ill-posedness of the problem. For the problems that are slightly ill-posed, we do not need a strong regularization, thus the optimal α need not to be too large. However,

[17]For example for α and γ_1 the coverage probabilities are lower under all selection criteria and for all sample sizes whereas for ρ_l selection using SIC leads to higher coverage probabilities.

Table 4: **Simulation Results with Diagonal Weighting Matrix, T=200**

	Bias, Standard Deviation, Coverage Probability								
	α_y			γ_1			ρ_l		
h	bias	std dev	prob	bias	std dev	prob	bias	std dev	prob
1	0.003	0.061	0.952	-0.001	0.074	0.947	0.003	0.096	0.957
2	-0.002	0.048	0.953	-0.002	0.053	0.948	-0.002	0.051	0.953
3	-0.001	0.046	0.950	-0.004	0.059	0.943	-0.003	0.042	0.956
4	-0.003	0.047	0.946	-0.006	0.056	0.938	-0.004	0.037	0.951
5	-0.003	0.048	0.948	-0.004	0.060	0.942	-0.004	0.034	0.953
6	-0.003	0.049	0.953	-0.005	0.059	0.941	-0.005	0.031	0.952
7	-0.003	0.050	0.957	-0.005	0.061	0.938	-0.005	0.030	0.957
8	-0.003	0.050	0.958	-0.006	0.060	0.944	-0.005	0.028	0.957
9	-0.004	0.051	0.959	-0.006	0.062	0.944	-0.005	0.027	0.959
10	-0.004	0.051	0.960	-0.006	0.061	0.945	-0.005	0.027	0.959
11	-0.003	0.051	0.961	-0.005	0.062	0.942	-0.005	0.026	0.961
12	-0.003	0.051	0.961	-0.006	0.062	0.943	-0.005	0.026	0.960

Simulation results for different number of horizons, h. Estimated parameters: Share of capital in production, α_y, AR(1) coefficient of the labour income tax process, ρ_l and the effect of change in labour productivity on employment, γ_1.
Source: Author's calculations

for severely ill-posed inverse problems, we need a strong regularization which means a larger α. In our case we expect the ill-posedness of the problem to increase with the horizon length. One way to measure ill-posedness is to compute the condition number of the covariance matrix of the IRFs. As defined in Greene (2012), condition number of a matrix is the ratio of the largest eigenvalue of a matrix to the smallest one. It may indicate the closeness of a matrix to singularity. Hence in our case, it indicates the degree of ill-posedness; the higher the condition number, the more ill-posed our problem is. We present the condition number for different values of h when $T = 200$ in Table 7. As can be seen, the ill-posedness of the problem increases with h. In Table 8, we present values of optimal α used in each simulation. For sample sizes $T = 100$ and $T = 400$ optimal α increases with horizon, h, as expected. However, α does not show a regular trend for $T = 200$ except for horizons 2 to 4 and 8 to 11. Finally, Figure 1 shows the objective function and the selected value of optimal α for each horizon for sample size 400. Note that for the selected grid, all of the objective functions except for $h = 1$ have a global minimum.

Table 5: **Simulation Results with Identity Weighting Matrix, T=200**

	Bias, Standard Deviation, Coverage Probability								
	α_y			γ_1			ρ_l		
h	bias	std dev	prob	bias	std dev	prob	bias	std dev	prob
1	0.003	0.061	0.949	-0.002	0.075	0.944	0.005	0.097	0.958
2	-0.002	0.049	0.953	-0.002	0.052	0.951	-0.002	0.051	0.952
3	-0.002	0.047	0.948	-0.005	0.063	0.946	-0.003	0.042	0.953
4	-0.003	0.048	0.944	-0.004	0.056	0.946	-0.004	0.037	0.946
5	-0.003	0.048	0.944	-0.006	0.062	0.945	-0.004	0.033	0.953
6	-0.003	0.049	0.941	-0.004	0.058	0.948	-0.005	0.031	0.954
7	-0.003	0.049	0.943	-0.006	0.061	0.943	-0.005	0.029	0.957
8	-0.003	0.049	0.942	-0.004	0.058	0.948	-0.005	0.028	0.958
9	-0.003	0.049	0.947	-0.006	0.060	0.945	-0.005	0.027	0.960
10	-0.003	0.050	0.945	-0.005	0.059	0.947	-0.005	0.026	0.961
11	-0.003	0.050	0.947	-0.006	0.060	0.946	-0.005	0.026	0.962
12	-0.003	0.050	0.945	-0.005	0.059	0.947	-0.005	0.026	0.961

Simulation results for different number of horizons, h. Estimated parameters: Share of capital in production, α_y, AR(1) coefficient of the labour income tax process, ρ_l and the effect of change in labour productivity on employment, γ_1.
Source: Author's calculations

6 Empirical Application

In this section we use the proposed regularized estimator to estimate the DSGE model in Altig et al. (2011). Altig et al. (2011) estimate a DSGE model so as to answer micro-macro pricing conflict, i.e. macroeconomic data suggesting inflation is inertial vs. microeconomic data showing that firms are changing their prices frequently. Under the assumption that capital is firm-specific, the authors are able to account for inflation inertia at the same time having firms re-optimizing their prices on average once every 1.5 quarters. They use an IRF matching estimator where they first estimate a VAR(4) model made of 10 key U.S. macroeconomic time series variables and they obtain the impulse responses of these variables to three shocks: a neutral technology shock, a capital embodied technology shock and a monetary policy shock. The parameter estimates are obtained by matching VAR impulse responses with those implied by the economic model. The IRF horizon is set to 20 and the weighting matrix is chosen to be a diagonal weighting matrix, i.e., a matrix with the inverse of the variances of the impulse responses on the diagonal and zero elsewhere. The data is quarterly and the sample period is 1959II-2001IV.

Table 6: **Simulation Results with Generalized Inverse Weighting Matrix, T=200**

Bias, Standard Deviation, Coverage Probability									
	α_y			γ_1			ρ_l		
h	bias	std dev	prob	bias	std dev	prob	bias	std dev	prob
1	0.005	0.060	0.955	0.010	0.070	0.952	0.000	0.091	0.954
2	-0.000	0.045	0.938	-0.000	0.038	0.932	-0.006	0.050	0.934
3	0.002	0.040	0.928	-0.001	0.034	0.929	-0.004	0.027	0.957
4	-0.002	0.046	0.810	0.001	0.037	0.848	-0.004	0.029	0.923
5	-0.004	0.047	0.740	0.001	0.039	0.796	-0.004	0.031	0.876
6	-0.003	0.046	0.724	0.000	0.039	0.776	-0.005	0.031	0.829
7	-0.002	0.045	0.715	-0.002	0.039	0.769	-0.005	0.032	0.752
8	-0.003	0.046	0.672	-0.002	0.040	0.746	-0.007	0.030	0.719
9	-0.003	0.048	0.651	-0.003	0.040	0.727	-0.008	0.030	0.681
10	-0.003	0.048	0.612	-0.002	0.041	0.704	-0.009	0.029	0.670
11	-0.002	0.049	0.592	-0.002	0.042	0.693	-0.010	0.029	0.634
12	-0.002	0.049	0.565	-0.002	0.041	0.688	-0.008	0.031	0.609

Simulation results for different number of horizons, h. Estimated parameters: Share of capital in production, α_y, AR(1) coefficient of the labour income tax process, ρ_l and the effect of change in labour productivity on employment, γ_1.
Source: Author's calculations

Table 7: **Condition number for different** h**,** $T = 200$

	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$
Condition #	4.43	12.98	94.08	1.143×10^3	3.80×10^3	1.45×10^4
	$h=7$	$h=8$	$h=9$	$h=10$	$h=11$	$h=12$
Condition #	4.87×10^4	2.86×10^5	1.02×10^6	5.01×10^6	1.53×10^7	5.99×10^7

Condition number denotes the ratio of the largest eigenvalue of the covariance matrix of the IRFs to the smallest one.
Source: Author's calculations

We estimate the same model by using the regularized weighting matrix and we pick the optimal regularization parameter using the rule given in Equation 7. The estimation is done by user written code in MATLAB where we modified the code used in Hall et al. (2012). The results are presented in Table 9. The first column of Table 9 gives the parameters.[18] Columns 2 and 3 of Table 9 present the results obtained by our

[18]For a brief description of the parameters, see Appendix D.

Table 8: **Optimal α's for different simulations**

	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$
$T = 100$	10^{-9}	2.89×10^{-6}	5.88×10^{-7}	4.92×10^{-6}	1.43×10^{-5}	1.43×10^{-5}
$T = 200$	10^{-9}	2.90×10^{-6}	4.92×10^{-6}	8.38×10^{-6}	10^{-9}	10^{-9}
$T = 400$	10^{-9}	1.70×10^{-6}	4.92×10^{-6}	2.89×10^{-6}	4.92×10^{-6}	4.92×10^{-6}
	$h = 7$	$h = 8$	$h = 9$	$h = 10$	$h = 11$	$h = 12$
$T = 100$	2.43×10^{-5}	4.13×10^{-5}	4.13×10^{-5}	4.13×10^{-5}	7.02×10^{-5}	1.19×10^{-9}
$T = 200$	10^{-9}	2.90×10^{-9}	4.92×10^{-9}	8.38×10^{-9}	8.38×10^{-9}	10^{-9}
$T = 400$	8.38×10^{-6}	8.38×10^{-6}	8.38×10^{-6}	8.38×10^{-6}	8.38×10^{-6}	8.38×10^{-6}

Source: Author's calculations

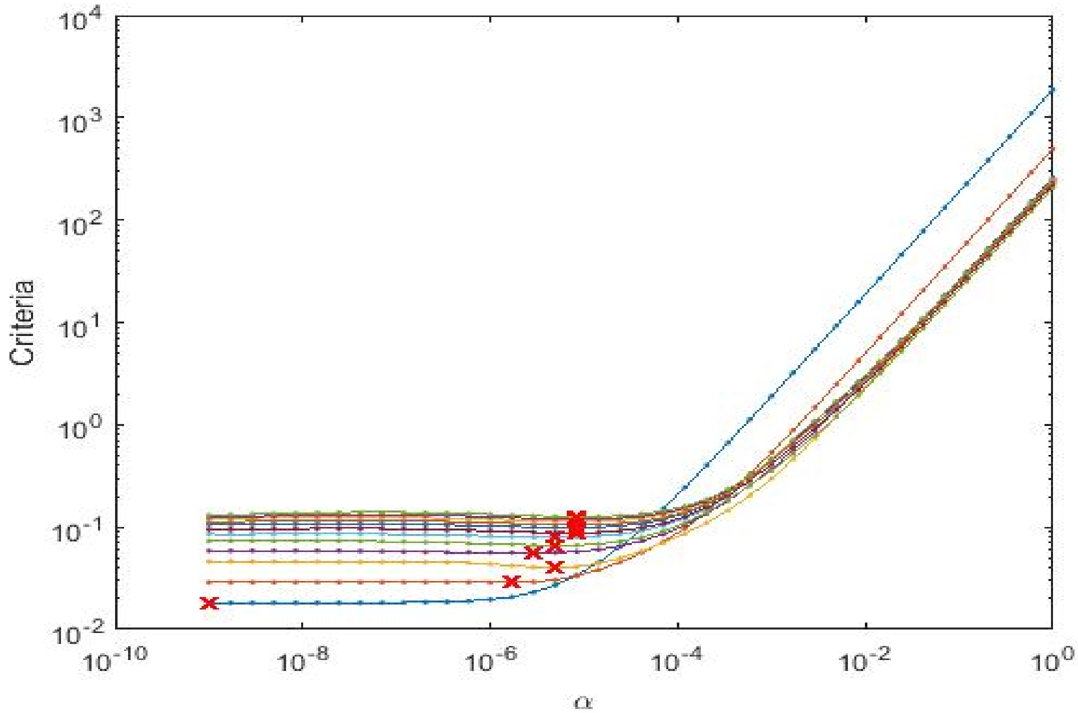

Figure 1: Criterion functions for α, $T = 400$. The curves from bottom to top are the criterion functions for $h = 1$ to $h = 12$, respectively. The crosses show the optimal α's. Source: Author's calculations

method whereas columns 4 and 5 show the results of Altig et al. (2011).[19],[20] For both

[19]We attribute the small differences between the results given in columns 4 and 5 and Altig et al. (2011) benchmark model to the change of the version of the MATLAB.

[20]Note that for the sake of exposition we briefly explain the model in Altig et al. (2011) in this section and report our results using the same notation as in Altig et al. (2011). For a complete description of the model we refer the reader to Altig et al. (2011).

of the estimators, we compute the standard errors using the sandwich formula given in equation (3).[21] We present the results where the standard errors of the regularized estimator are computed by using the optimal variance formula in Appendix D, Table 10.

Table 9: Estimation Results of the model in Altig et al. (2011)

	Regularized		Diagonal	
Parameters	Parameter Estimates	Standard Errors	Parameter Estimates	Standard Errors
ρ_{xM}	-0.048	0.288	-0.034	0.294
ρ_{xz}	0.309	0.839	0.328	0.948
c_z	2.604	3.013	2.997	3.010
ρ_{μ_z}	0.814	0.188	0.904	0.140
$\rho_{x\Upsilon}$	0.781	0.644	0.822	0.349
c_Υ	-0.008	0.409	0.246	0.442
$\rho_{\mu\Upsilon}$	0.236	1.075	0.241	0.425
σ_M	0.314	0.094	0.329	0.109
σ_{μ_z}	0.059	0.055	0.067	0.062
$\sigma_{\mu\Upsilon}$	0.271	0.336	0.303	0.093
ϵ	0.951	0.618	0.800	0.255
S''	3.345	3.025	3.276	3.413
ξ_w	0.717	0.167	0.723	0.261
b	0.708	0.151	0.705	0.135
σ_a	2.091	2.209	2.018	4.172
c_z^p	1.130	3.551	1.421	3.779
c_Υ^p	-0.109	0.542	0.134	0.499
γ	0.038	0.047	0.040	0.071
Implied Average Time Between Re-Optimization				
Firm-Specific Capital Model	1.527	0.005	1.508	0.067
Homogeneous Capital Model	5.771	0.033	5.603	0.455
Number of observations	167		167	
Number of horizons	20		20	
Norm of the standard errors	6.2554		7.3705	

Source: Author's calculations

First of all, the parameter estimates from two different estimations are very close

[21] Sandwich formula is often used to refer the asymptotic variance estimator of the GMM estimator, see Ruud (2000), page 549.

to each other. Second, the standard errors for regularized estimator are mostly smaller as it can also be seen from the smaller norm. Third, statistical significance of 3 out of 18 parameters is found to be different under the two estimation methods. The first of these parameters is $\rho_{x\Upsilon}$. $\rho_{x\Upsilon}$ is the AR coefficient of $\hat{x}_{\Upsilon,t}$ which is the response of monetary policy to an innovation in capital embodied technology and this response is assumed to be characterized by an $ARMA(1,1)$ process. Different from Altig et al. (2011), we found $\rho_{x\Upsilon}$ to be insignificant, hence our results indicate that $\hat{x}_{\Upsilon,t}$ might indeed be characterized by an $MA(1)$ process implying that the monetary policy response to capital embodied technology does not depend on responses in the previous period, so the effects of the shocks are short-term. In other words, there is no persistence in the response of monetary policy to technology shocks. Note that, this is also the case for the monetary policy response to neutral technology shock, $\hat{x}_{z,t}$ as its AR coefficient, ρ_{xz}, is found to be insignificant in both estimations using regularized and diagonal weighting matrices. Second parameter estimated to be insignificant is $\sigma_{\mu\Upsilon}$. Υ_t denotes capital embodied technology which affects the fixed cost of production of intermediate goods. The parameter $\sigma_{\mu\Upsilon}$ shows the short-run effect of a shock to capital embodied technology on Υ_t. According to the estimates obtained with a diagonal matrix, one standard deviation shock to embodied technology drives Υ_t up by 0.30 percent. According to our results, this effect is smaller, 0.271 percent however it is not significantly different from zero, meaning there is no short-run effect of a capital embodied technology shock. Finally, ϵ is defined as the interest semi-elasticity of money demand in steady state. The point estimate in Altig et al. (2011) is 0.8 and significant, implying that one percentage increase in annualized rate of interest results in almost one percent decline in real transactions. Our point estimate is 0.951 and its p-value is just above 10%, it is slightly insignificant. An insignificant ϵ implies that changes in annualized rate of interest have no significant effect on real transactions which is unlikely to hold. So, this result may stem from the lack of identifying power of the data.[22]

To sum up, the application exercise in this section shows that using a regularized inverse does not change the point estimates drastically, however the standard errors might be smaller than those obtained with a diagonal weighting matrix. Moreover, as proved in Section 4, we know that the asymptotic distribution of the regularized

[22] In Appendix D, Table 10, we also report estimation results obtained with the generalized inverse of the optimal weighting matrix. None of the parameters are found to be significant.

estimator converges to the that of the optimal estimator which will give a J-statistic asymptotically distributed as χ^2 with degrees of freedom implied by the model.[23]

7 Conclusion

In this paper, we propose to use regularization to address the invertibility problem of the efficient weighting matrix that arises in IRFME. After defining the regularized estimator, we establish its asymptotic properties and analyse its small sample performance with Monte Carlo simulations. We then use it to estimate the benchmark model in Altig et al. (2011).

Contributions of the paper are as follows: First, we are the first to use *Tikhonov regularization* in IRFME. Second, we show that the estimation using a regularized weighting matrix performs best compared to the estimation obtained with a generalized inverse. Although the regularized estimator performs slightly worse than the estimators with diagonal and identity weighting matrices in terms of coverage probability, it in general has smaller bias and variance. Third, we show that for sufficiently large samples, regularized estimation is less affected by the choice of the number of impulse response horizons compared to the estimator with generalized inverse, i.e., the regularization parameter adapts itself such that increasing the length of the horizon does not deteriorate the performance of the estimator too much, especially when the sample size is large. Thus, unlike existing approaches, selection of the optimal horizon is not a critical issue with regularized estimation.

Selection of the regularization parameter is a crucial issue in the estimation. In our application example, we use a data driven rule but we do not prove its optimality theoretically. Theoretical work on the selection of optimal regularization parameter in regularized IRFME is left for future work.

[23] We do not report J-statistics coming from the two different estimations in Table 9, as the J-statistic obtained with two different matrices are converging to two different limiting distributions, hence they are not comparable. We found that J-statistic obtained with the regularized inverse is equal to 1.1571×10^6 and it results in rejection of the model. Note that Feve et al. (2009) also reject the model that they estimate when they use the bootstrap distribution of the J-statistic instead of the standard limiting distribution.

Appendices

A Technical Proofs

A.1 Proof of Lemma 1

Proof. Proof of Lemma 1 follows from the proof of Theorem 7 in Carrasco and Florens (2000). Nonetheless below we present the proof in our case step by step to make it clear for the reader. Let us write:

$$W^{\alpha} f = (\alpha I + W'W)^{-1} W' f$$

$$\hat{W}^{\alpha} \hat{f} = (\alpha I + \hat{W}'\hat{W})^{-1} \hat{W}' \hat{f}$$

where we drop the arguments on which f and $\hat{f}$ functions depend for the sake of exposition. Then

$$\left\| \hat{W}^{\alpha} \hat{f} - W^{+} f \right\|^2 \leq \underbrace{\left\| \hat{W}^{\alpha} \hat{f} - W^{\alpha} f \right\|^2}_{I} + \underbrace{\left\| W^{\alpha} f - W^{+} f \right\|^2}_{II}$$

To prove the Lemma 1, we need to show that the first and the second term on the right hand side converges to zero. Let us begin by the first term:

$$\left\| \hat{W}^{\alpha} \hat{f} - W^{\alpha} f \right\|^2 = \left\| (\alpha I + \hat{W}'\hat{W})^{-1} \hat{W}' \hat{f} - (\alpha I + W'W)^{-1} W' f \right\|^2$$

$$= \left\| (\alpha I + \hat{W}'\hat{W})^{-1} (\hat{W}' \hat{f} - W' f) + (\alpha I + \hat{W}'\hat{W})^{-1} W' f - (\alpha I + W'W)^{-1} W' f \right\|^2$$

$$\leq \left\| (\alpha I + \hat{W}'\hat{W})^{-1} \right\|^2 \left\| (\hat{W}' \hat{f} - W' f) \right\|^2 + \left\| (\alpha I + \hat{W}'\hat{W})^{-1} W' f - (\alpha I + W'W)^{-1} W' f \right\|^2$$

The first term in (I) is $O(\frac{1}{\alpha^2})$ by Proposition 3.2 of Darolles et al. (2011), while the second term is $O(\frac{1}{T})$. The third term needs a further investigation:

$$\left\| (\alpha I + \hat{W}'\hat{W})^{-1} W' f - (\alpha I + W'W)^{-1} W' f \right\|^2$$

$$\leq \left\| (\alpha I + \hat{W}'\hat{W})^{-1} W' f - (\alpha I + \hat{W}'\hat{W})^{-1} W'W (\alpha I + W'W)^{-1} W' f \right\|^2 \qquad (1)$$

$$+\left\|(\alpha I+\hat{W}'\hat{W})^{-1}\hat{W}'\hat{W}(\alpha I+W'W)^{-1}W'f-(\alpha I+W'W)^{-1}W'f\right\|^2 \qquad (2)$$

$$+\left\|(\alpha I+\hat{W}'\hat{W})^{-1}W'W(\alpha I+W'W)^{-1}W'f-(\alpha I+\hat{W}'\hat{W})^{-1}\hat{W}'\hat{W}(\alpha I+W'W)^{-1}W'f\right\|^2 \qquad (3)$$

Let us investigate the third term of (I) again term by term.

$$(1)\leq\left\|(\alpha I+\hat{W}'\hat{W})^{-1}W'W\right\|^2\left\|(W^{-1}-(\alpha I+W'W)^{-1}W')f\right\|^2$$

$$=\underbrace{\left\|(\alpha I+\hat{W}'\hat{W})^{-1}W'W\right\|^2}_{\leq 1}\underbrace{\left\|(W^{-1}-W^{\alpha})f\right\|^2}_{O(\alpha^2)}$$

rates follow from proof of Theorem 7 in Kress (1999) and Carrasco and Florens (2000).

$$(2)\leq\left\|(\hat{W}^{\alpha}-\hat{W}^{-1})\hat{W}(\alpha I+W'W)^{-1}W'f\right\|^2$$

$$\leq\underbrace{\left\|\hat{W}(\alpha I+W'W)^{-1}W'\right\|^2}_{\leq 1}\underbrace{\left\|(\hat{W}^{\alpha}-\hat{W}^{-1})f\right\|^2}_{O(\alpha^2)}$$

again, rates follow from proof of Theorem 7 in Carrasco and Florens (2000) and Kress (1999), respectively.

$$(3)\leq\left\|(\alpha I+\hat{W}'\hat{W})^{-1}\right\|^2\left\|W'W-\hat{W}'\hat{W}\right\|^2\left\|W^{\alpha}f\right\|^2$$

The first term is $O(\frac{1}{\alpha^2})$ by Proposition 3.2 in Darolles et al. (2011), the second term is $O(\frac{1}{T})$ and finally the third term is $O(\frac{1}{\alpha})$ by Theorem 7 Carrasco and Florens (2000). So (I) converges to zero as $T\to\infty$, $\alpha^3 T\to\infty$ and $\alpha\to 0$.

To show that (II) converges to 0, we use Fourier decompositions:

$$W^{+}f=\sum_{j=1}^{r}\frac{1}{\lambda_j}\langle f,\phi_j\rangle\,\phi_j$$

$$W^{\alpha}f=\sum_{j=1}^{r}\frac{\lambda_j}{\alpha+\lambda_j^2}\langle f,\phi_j\rangle\,\phi_j$$

where λ_j and ϕ_j are the eigenvalues and corresponding eigenfunctions of the matrix

W, respectively and r is its rank. Then:

$$\left\|W^\alpha f - W^+ f\right\|^2 = \sum_{j=1}^{r} \left(\frac{\lambda_j}{\alpha + \lambda_j^2} - \frac{1}{\lambda_j} \right)^2 \langle f, \phi_j \rangle^2$$

$$= \sum_{j=1}^{r} \left(\frac{-\alpha}{(\alpha + \lambda_j^2)\lambda_j} \right)^2 \langle f, \phi_j \rangle^2$$

$$= \sum_{j=1}^{r} \frac{\alpha^2}{\left((\alpha + \lambda_j^2)\lambda_j\right)^2} \langle f, \phi_j \rangle^2$$

$$= \alpha^2 \sum_{j=1}^{r} \frac{1}{\left((\alpha + \lambda_j^2)\lambda_j\right)^2} \langle f, \phi_j \rangle^2 \leq \alpha^2 \sum_{j=1}^{r} \frac{1}{(\lambda_j^3)^2} \langle f, \phi_j \rangle^2$$

So (II) is $O(\alpha^2)$. Then we can conclude that:

$$\left\|\hat{W}^\alpha \hat{f} - W^+ f\right\| \to 0 \quad as \quad \alpha \to 0, T \to \infty \quad \text{and} \quad \alpha^3 T \to \infty$$

■

A.2 Proof of Theorem 2

Proof. The proof follows in three steps:

1. First, we show the distribution of the estimator that uses pseudo-inverse as the weighting matrix.

2. We show that following step 1 of Carrasco and Florens (2000), the estimator with regularized inverse converges in distribution to the one with the generalized inverse.

3. We show that the estimator we obtain in Step 1 is the optimal one.

1st Step
Let $\hat{\theta}^{+}$ be the estimator given by the following minimization problem:

$$\hat{\theta}^{+} = \underset{\theta}{\operatorname{argmin}}\, f'(\theta) A_T' A_T f(\theta)$$

where A_T is a sequence of linear operators, $A_T' A_T$ converging to W^{+}, the pseudo-inverse of the variance-covariance matrix of IRFs.

Following Gourieroux et al. (1995), one can then show that:

$$\sqrt{n}(\hat{\theta}^{+} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

where

$$\Sigma = \left(\frac{\partial f'}{\partial \theta} W^{+} \frac{\partial f}{\partial \theta'} \right)^{-1}$$

2nd Step

We need to show that:

$$\frac{\partial \hat{f}'}{\partial \theta}(\hat{\theta}) \hat{W}^{\alpha} \sqrt{n} \hat{f}(\theta_0) \xrightarrow{d} \frac{\partial f'}{\partial \theta}(\theta_0) W^{+} f(\theta_0)$$

Let us write:

$$\frac{\partial \hat{f}'}{\partial \theta}(\hat{\theta}) \hat{W}^{\alpha} \sqrt{n} \hat{f}(\theta_0) = \underbrace{\left(\frac{\partial \hat{f}'}{\partial \theta}(\hat{\theta}) \hat{W}^{\alpha} - \frac{\partial f'}{\partial \theta}(\theta_0) W^{+} \right) \sqrt{n} \hat{f}(\theta_0)}_{I}$$

$$+\underbrace{\frac{\partial f'}{\partial \theta}(\theta_0)W^{+}\sqrt{n}\hat{f}(\theta_0)}_{II}$$

$$I \leq \underbrace{\|\frac{\partial \hat{f}'}{\partial \theta}(\hat{\theta})\hat{W}^{\alpha} - \frac{\partial f'}{\partial \theta}(\theta_0)W^{+}\|}_{o_p(1)} + \underbrace{\|\sqrt{n}\hat{f}(\theta_0)\|}_{O_p(1)}$$

by Lemma 1. Moreover it is straightforward to show that:

$$II \xrightarrow{d} \frac{\partial f'}{\partial \theta}(\theta_0)W^{+}f(\theta_0)$$

Then it follows that:

$$\sqrt{n}(\hat{\theta}^{\alpha} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

where $\Sigma = \left(\frac{\partial f'}{\partial \theta}W^{+}\frac{\partial f}{\partial \theta'}\right)^{-1}$.

3rd Step

In this step we need to show that the estimator obtained with W^{+} is the optimal one. Let A a symmetric positive definite weighting matrix. The variance of the estimator obtained with A is given by:

$$Var(\hat{\theta}_A) = \left(\frac{\partial f'}{\partial \theta}A\frac{\partial f}{\partial \theta'}\right)^{-1}\frac{\partial f'}{\partial \theta}AWA\frac{\partial f}{\partial \theta'}\left(\frac{\partial f'}{\partial \theta}A\frac{\partial f}{\partial \theta'}\right)^{-1}$$

Then we need to show that:

$$S = Var(\hat{\theta}_A) - \Sigma \geq 0$$

$$\left(\frac{\partial f'}{\partial \theta}A\frac{\partial f}{\partial \theta'}\right)^{-1}\frac{\partial f'}{\partial \theta}AWA\frac{\partial f}{\partial \theta'}\left(\frac{\partial f'}{\partial \theta}A\frac{\partial f}{\partial \theta'}\right)^{-1} - \left(\frac{\partial f'}{\partial \theta}W^{+}\frac{\partial f}{\partial \theta'}\right)^{-1}$$

Let $B = \left(\frac{\partial f'}{\partial \theta}A\frac{\partial f}{\partial \theta'}\right)^{-1}$. Then, after some matrix algebra manipulation, one can write:

$$S = B\frac{\partial f'}{\partial \theta}A\left[W - \frac{\partial f}{\partial \theta}\left(\frac{\partial f'}{\partial \theta}W^{+}\frac{\partial f}{\partial \theta'}\right)^{-1}\frac{\partial f'}{\partial \theta}\right]A\frac{\partial f}{\partial \theta'}B$$

Since W is positive semidefinite, it has a Cholesky decomposition. Hence one can write $W = D'D$. Moreover by Theorem 5 in Chapter 2, Section 7 in Magnus and

Neudecker (1988): $(D'D)^+ = D^+(D^+)'$. Hence:

$$S = B\frac{\partial f'}{\partial \theta} AD' \left[I - (D')^+ \frac{\partial f}{\partial \theta} \left(\frac{\partial f'}{\partial \theta} D^+(D^+)' \frac{\partial f}{\partial \theta'} \right)^{-1} \frac{\partial f'}{\partial \theta} D^+ \right] DA \frac{\partial f}{\partial \theta'} B$$

Since the term in the middle is an orthogonal projection matrix, S is positive semi-definite.

■

B Extension of the results to the case $h = \infty$

In Section 4, we presented the results for $h < \infty$. However, our method allows for $h = \infty$. In this section, we extend our results to the case where $h = \infty$.

Let us redefine the function $f(\infty, \theta) \in \mathcal{E}$, where $\mathcal{E}$ is a Hilbert space:

$$f(\infty, \theta_0) = \varphi_\infty - \psi_\infty(\theta_0) \tag{8}$$

where $\varphi_\infty \in \mathcal{E}$ is the infinite dimensional vector of IRFs coming from the structural VAR and depends on the process $\{X_t\}$, $X_t \in \mathbb{R}^{d_X}$. $\psi_\infty(\theta_0) \in \mathcal{E}$ is the infinite dimensional vector of IRFs implied by the structural DSGE model where θ_0 is the true value of vector of parameters. Let A be a linear operator defined on $\mathcal{E}$. Note that, since $h = \infty$, A is no longer finite dimensional. The estimate of θ_0 is given by:

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \|A_T \hat{f}(\infty, \theta)\|^2$$

where A_T is a sequence of linear operators converging to A and $\hat{f}$ is the empirical counterpart of (8):

$$\hat{f}(\infty, \theta) = \hat{\varphi}_\infty - \psi_\infty(\theta)$$

Some of the assumptions stated in Section 4 are needed to be modified for the case $h = \infty$:

Assumption 2. 1 *(i) Let φ_∞ be the IRFs up to horizon $h = \infty$ from the reduced form VAR. $\varphi_\infty : \mathbb{R}^r \to \mathcal{E}$, where r is the number of existing moments of $\{X_t\}_{t=1}^T$ and $\mathcal{E}$ is a Hilbert space endowed with the scalar product $\langle g(z), h(z) \rangle = \int_Z g(z)h(z)f(z)dz$ and hence the norm in $\mathcal{E}$ is given by $\|g\|^2 = \langle g, g \rangle$.*

(ii) Let $\psi_\infty(\theta)$ be a function of theoretical IRFs from the structural DSGE model up to horizon $h = \infty$. $\psi_\infty : \mathbb{R}^k \to \mathcal{E}$, where $\theta \in \Theta \in \mathbb{R}^k$ is the vector of parameters to be estimated.

Assumption 4. 1 *Let W^{-1} is the Moore-Penrose inverse of the covariance operator of φ_∞. Then $f(\infty, \theta) \in \mathcal{H}(W^{-1}) + \mathcal{H}(W^{-1})^\perp$ for all θ where $\mathcal{H}(W^{-1})$ is a reproducing kernel Hilbert space and $\mathcal{H}(W^{-1})^\perp$ is its orthogonal complement.*

Assumption 6. 1 *$f(\infty, \theta)$ is differentiable with respect to θ.*

Assumption 9. 1 *The $(k \times k)$ matrix $\langle A_T \frac{\partial \hat{f}(\infty,\hat{\theta})}{\partial \theta'}, A_T \frac{\partial \hat{f}(\infty,\hat{\theta})}{\partial \theta'} \rangle$ is positive definite and symmetric.*

Since $h = \infty$, the proof involves working with infinite dimensional vectors, we need to state some further assumptions to show the asymptotic normality.

Assumption 13 *The inner product satisfies the following differentiation rule:*

$$\frac{\partial}{\partial \theta'} \langle u(\theta), v(\theta) \rangle = \left\langle \frac{\partial}{\partial \theta'} u(\theta), v(\theta) \right\rangle + \left\langle u(\theta), \frac{\partial}{\partial \theta'} v(\theta) \right\rangle$$

A and A_T commute with differential operator:

$$\frac{\partial}{\partial \theta'}[Au(\theta)] = A[\frac{\partial}{\partial \theta'} u(\theta)]$$

Assumption 14 *$\sqrt{T}\hat{\varphi}_\infty$ converges in law to Y as $T \to \infty$, where $Y \sim \mathcal{N}(0, W) \in \mathcal{E}$.*

Assumption 15 *Consider the covariance operator with kernel: $a(t,s) = E(f_t(\infty,\theta) f_s(\infty,\theta))$. Then the covariance kernel $a(t,s)$ is an L^2 kernel.*

Lemma 4 *Let $W^\alpha = (\alpha I + W^* W)^{-1} W^*$ be the Tikhonov regularized inverse of the operator W such that $A^* A = W^{-1}$. Assume that Assumption 4.1 holds. Moreover let $\hat{W}^\alpha = (\alpha I + \hat{W}^* \hat{W})^{-1} \hat{W}^*$ be the regularized inverse $\hat{W}$ such that $A_T' A_T = \hat{W}^{-1}$ and $\hat{W}$ converges to W, Then:*

$$\left\| \hat{W}^\alpha \hat{f}(\infty, \theta) - W^{-1} f(\infty, \theta) \right\| \to 0 \quad \text{in probability as} \quad T \to \infty, \alpha^3 T \to \infty, \alpha \to 0$$

Proof. The proof follows from the proof of Theorem 7 of Carrasco and Florens (2000). ■

Theorem 5 *Under the Assumptions 1,2.1,3,4.1,5,6.1,7,8,9.1,11,12,13, 14 and 15, the estimator:*

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \left\| \hat{f}(\infty, \theta) \right\|^2_{\hat{W}^{\alpha}}$$

satisfies:
(i) $\hat{\theta} \to \theta_0$ *in probability as* $T \to \infty$, $\alpha^3 T \to \infty$ *and* $\alpha \to 0$
(ii) $\sqrt{T}(\hat{\theta} \to \theta_0) \xrightarrow{d} \mathcal{N}\left(0, \left\langle W^{-1} \frac{\partial f}{\partial \theta'}(\infty, \theta_0), \frac{\partial f}{\partial \theta'}(\infty, \theta_0) \right\rangle^{-1}\right)$ *as* $T \to \infty$, $\alpha^3 T \to \infty$ *and* $\alpha \to 0$

Proof. The proof of Theorem 5 is also done in three steps and it follows the proof of Theorem 8 in Carrasco and Florens (2000) closely. In the first step, the asymptotic distribution of the estimator for any A is obtained following the proof of Theorem 2 in Carrasco and Florens (2000). In the second step the asymptotic distribution obtained with regularized inverse is shown to converge to the distribution of the estimator obtained with generalized inverse. Finally in the third step, it is shown that the estimator obtained with the generalized inverse is the optimal one. Steps two and three follow from the proof of Theorem 8 in Carrasco and Florens (2000). ■

C RBC Model in Christiano et al. (2006)

In this section, we briefly present the model in Christiano et al. (2006) on which the simulation design in based. As mentioned in Section 5, it is a simple RBC model where the only shock which effects labour productivity in the long run is a shock to technology.

In the model, a representative agent chooses his per capita consumption c_t and per capita hours worked, l_t, to maximize his expected utility given his budget constraint:

$$\max E_0 \sum_{t=0}^{\infty} (\beta(1+\gamma))^t \left[log c_t + \psi \frac{(1-l_t)^{1-\sigma} - 1}{1-\sigma} \right]$$

$$\text{subject to} \qquad c_t + (1+\tau_{x,t}) i_t \leq (1 - \tau l, t) w_t l_t + r_t k_t + T_t$$

where $i_t = (1+\gamma)k_{t+1} - (1-\delta)k_t$. $\tau_{x,t}$ is the investment tax, $\tau_{l,t}$ is the tax rate on labour income, w_t is the wage rate, r_t is the rental rate on capital and k_t is the per capita capital stock. Moreover T_t denotes lump sum taxes and $\delta \in (0,1)$ is the capital's depreciation rate while γ is the population growth rate. Finally, $\sigma > 0$ is a curvature parameter.

On the firms' side, production function of the representative firm is given by:

$$y_t = k_t^\alpha (Z_t l_t)^{1-\alpha}$$

where Z_t is the state of the technology at time t and $\alpha \in (0,1)$. Christiano et al. (2006) then define the stochastic processes for the shocks as the following:

$$log z_t = \mu_z + \sigma_z \varepsilon_t^z$$

$$\tau_{l,t+1} = (1-\rho_l)\tau_l + \rho_l \tau_{l,t} + \sigma_l \varepsilon_{t+1}^l$$

$$\tau_{x,t+1} = (1-\rho_x)\tau_x + \rho_x \tau_{x,t} + \sigma_x \varepsilon_{t+1}^x$$

where $z_t = Z_t/Z_{t-1}$, μ_z is the mean growth rate of the technology, τ_l is the mean tax rate on labour income and τ_x is the mean capital tax. Moreover ε_t^z, ε_{t+1}^l and ε_{t+1}^x are i.i.d. random variables with mean 0 and variance equals to 1. The parameters, σ_z,σ_l and σ_x are nonnegative scalars. Moreover, the shocks are assumed to be stationary, i.e., ρ_l and ρ_x are assumed to be less than unity in absolute value.

The final equation of Christiano et al. (2006)'s RBC model is given by the resource constraint:

$$c_t + (1+\gamma)k_{t+1} - (1-\delta)k_t \leq y_t$$

D Empirical Application

D.1 Parameter Descriptions

Empirical application presented in Section 6 re-estimates the DSGE model in Altig et al. (2011). Altig et al. (2011) is a very large scale DSGE model where a VAR model of 10 key US macroeconomic variables is estimated and impulse responses to the three shocks are obtained: neutral technology shock, capital embodied technology shock and monetary policy shock.

For the sake of exposition we do not summarize Altig et al. (2011) model in this paper, however we briefly describe the parameters presented in Tables 9 and 10 for the ease of the interpretation of our results.

Altig et al. (2011) divide the parameters of their model into 3 groups: ζ_1, ζ_2,

ζ_3. The first group ζ_1 contains parameters such as discount factor and depreciation rate, which are calibrated. Second and third group parameters are estimated using IRFME. Hence, these are the parameters which appear in Tables 9 and 10.

Second group parameters are those which belong to the non-stochastic part of the model:

$$\zeta_2 = [\lambda_f, \xi_w, \gamma, \sigma_a, b, S'', \epsilon]$$

.

These parameters are:

- λ_f: CES production function parameter.[24]
- ξ_w: Proportion of the households who are able to optimize their nominal wage.
- γ: Coefficient of the marginal cost in the new Keynesian Phillips curve.
- σ_a: Inverse elasticity of capital utilization with respect to rental rate of capital (curvature of the cost function of utilization rate of capital).
- b: Habit parameter of households.
- S'': Inverse elasticity of investment with respect to price of installed capital (curvature of the S function in investment adjustment cost).
- ϵ: Interest semi-elasticity of money demand in steady state.

Third group parameters, ζ_3, belong to the stochastic part of the model. They capture the evolution of shocks as well as response of monetary policy to these shocks:

$$\zeta_3 = [\rho_{xM}, \sigma_M, \rho_{\mu_z}, \sigma\mu_z, \rho_{xz}, c_z, c_z^p, \rho_{\mu\Upsilon}, \sigma_{\mu\Upsilon}, \rho_{x}\Upsilon, c_\Upsilon, c_\Upsilon^p]$$

The first two parameters in ζ_3 characterize monetary policy shock. The next 5 parameters characterize the natural technology shock and response of monetary policy to this shock while the last 5 parameters characterize capital embodied technology shock and response of monetary policy to it. For more details, see equations (2.7), (2.8) and (2.19) in Altig et al. (2011).

[24]The authors fix the value of parameter λ_f to 1.05 for their main estimation results. For this reason, we also fix it to the same value and it does not appear in our results table.

- ρ_{xM}: AR(1) coefficient of monetary policy shock.
- σ_M: Standard deviation of monetary policy shock.
- ρ_{μ_z}: AR(1) coefficient on the growth rate of neutral technology shock.
- $\sigma\mu_z$: Standard deviation of the innovation in growth rate of neutral technology shock (short run effect of a shock to neutral technology shock).
- ρ_{xz}: AR(1) coefficient on the monetary policy response to neutral technology shock.
- c_z: Coefficient on the time t innovation in neutral technology in the $ARMA(1,1)$ process of monetary policy response to neutral technology shocks.
- c_z^p: Coefficient on the time $(t-1)$ innovation in neutral technology in the $ARMA(1,1)$ process of monetary policy response to neutral technology shocks.
- ρ_{μ_Υ}: AR(1) coefficient on the growth rate of capital embodied technology shock.
- σ_{μ_Υ}: Standard deviation of the innovation in growth rate of capital embodied technology shock (short run effect of a shock to capital embodied technology shock).
- $\rho_{x\Upsilon}$: AR(1) coefficient on the monetary policy response to capital embodied technology shock.
- c_Υ: Coefficient on the time t innovation in capital embodied technology in the $ARMA(1,1)$ process of monetary policy response to capital embodied technology shocks.
- c_Υ^p: Coefficient on the time $(t-1)$ innovation in capital embodied technology in the $ARMA(1,1)$ process of monetary policy response to capital embodied technology shocks.

D.2 Results

Below we present the results from two different estimations. Columns 2 and 3 shows the results where the standard errors of the regularized estimator are computed using optimal variance formula. The qualitative results are still the same. The significance of 3 out of 18 parameters is found to be different from the results obtained with diagonal weighting matrix. Columns 4 and 5 shows the estimation results where the generalized inverse of the covariance matrix of IRFs is used as the weighting matrix in IRFME. With the generalized inverse none of the parameters are significant and some of the standard errors are very large.

Table 10: Estimation Results of the model in Altig et al. (2011)

	Regularized		Generalized	
Parameters	Parameter Estimates	Standard Errors	Parameter Estimates	Standard Errors
ρ_{xM}	-0.048	0.338	-0.360	10.235
ρ_{xz}	0.309	0.938	0.002	1.513
c_z	2.604	3.747	0.327	2.577
ρ_{μ_z}	0.814	0.241	-0.029	1.033
$\rho_{x\Upsilon}$	0.781	0.799	0.803	960.422
c_Υ	-0.008	0.491	-1.163	20030.244
$\rho_{\mu\Upsilon}$	0.236	1.728	0.167	4514.953
σ_M	0.314	0.106	0.002	0.016
σ_{μ_z}	0.059	0.067	0.013	0.019
$\sigma_{\mu\Upsilon}$	0.271	0.565	0.000	0.0133
ε	0.951	1.080	3.512	11.335
S''	3.345	3.530	5.944	15.302
ξ_w	0.717	0.214	0.749	0.439
b	0.708	0.243	0.776	0.404
σ_a	2.091	2.690	3.522	22.294
c_z^p	1.130	3.840	1.504	2.68
c_Υ^p	-0.109	0.684	-1.259	22123.378
γ	0.038	0.064	0.074	0.357
Implied Average Time Between Re-Optimization				
Firm-Specific Capital Model	1.527	0.005	1.267	0.011
Homogeneous Capital Model	5.771	0.033	4.254	0.090
Number of observations	167		167	
Number of horizons	20		20	

Source: Author's calculations

References

Altig, D., L. Christiano, M. Eichenbaum, and J. Linde (2011): "Firm-Specific Capital, Nominal Rigidities and the Business Cycle," *Review of Economic Dynamics*, 14, 225–247.

Altug, S. (1989): "Time-to-build and aggregate fluctuations: some new evidence," *International Economic Review*, 30, 889–920.

An, S. and F. Schorfheide (2007): "Bayesian analysis of DSGE models," *Econometric reviews*, 26, 113–172.

Bekker, P. A. (1994): "Alternative Approximations to the Distributions of Instrumental Variable Estimators," *Econometrica*, 62, 657–81.

Carrasco, M. (2012): "A regularization approach to the many instruments problem," *Journal of Econometrics*, 170, 383–398.

Carrasco, M., M. Chernov, J.-P. Florens, and E. Ghysels (2007a): "Efficient estimation of general dynamic models with a continuum of moment conditions," *Journal of Econometrics*, 140, 529–573.

Carrasco, M. and J.-P. Florens (2000): "Generalization of GMM to a Continuum of Moment Conditions," *Econometric Theory*, 16, 797–834.

——— (2014): "On the Asymptotic Efficiency of GMM," *Econometric Theory*, 30, 372–406.

Carrasco, M., J.-P. Florens, and E. Renault (2007b): "Linear Inverse Problems in Structural Econometrics Estimation Based on Spectral Decomposition and Regularization," in *Handbook of Econometrics*, ed. by J. Heckman and E. Leamer, Elsevier, vol. 6 of *Handbook of Econometrics*, chap. 77.

Christiano, L. J. (1988): "Why does inventory investment fluctuate so much?" *Journal of Monetary Economics*, 21, 247–280.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1–45.

Christiano, L. J., M. Eichenbaum, R. Vigfusson, P. J. Kehoe, and M. W. Watson (2006): "Assessing Structural VARs [with Comments and Discussion]," *NBER Macroeconomics Annual*, 21, 1–105.

Darolles, S., Y. Fan, J. P. Florens, and E. Renault (2011): "Nonparametric Instrumental Regression," *Econometrica*, 79, 1541–1565.

Donald, S. G., G. W. Imbens, and W. K. Newey (2003): "Empirical likelihood estimation and consistent tests with conditional moment restrictions," *Journal of Econometrics*, 117, 55–93.

Donald, S. G. and W. K. Newey (2001): "Choosing the number of instruments," *Econometrica*, 69, 1161–1191.

Dovonon, P. and S. Gonçalves (2014): "Bootstrapping the GMM overidentification test under first-order underidentification," Working paper, Universite de Montreal, Montreal, QC.

Dovonon, P. and E. Renault (2013): "Testing for Common Conditionally Heteroskedastic Factors," *Econometrica*, 81, 2561–2586.

Dridi, R., A. Guay, and E. Renault (2007): "Indirect inference and calibration of dynamic stochastic general equilibrium models," *Journal of Econometrics*, 136, 397–430.

Engl, H. W., M. Hanke, and A. Neubauer (1996): *Regularization of Inverse Problems*, Dordrecht, Netherlands: Kluwer Academic Publications.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, and F. Schorfheide (2016): "Solution and estimation methods for DSGE models," in *Handbook of Macroeconomics*, Elsevier, vol. 2, 527–724.

Feve, P., J. Matheron, and J.-G. Sahuc (2009): "Minimum Distance Estimation and Testing of DSGE Models from Structural VARs," *Oxford Bulletin of Economics and Statistics*, 71, 883–894.

Florens, J.-P. and S. Sokullu (2017): "Nonparametric estimation of semiparametric transformation models," *Econometric Theory*, 33, 839–873.

Giannoni, M. and M. Woodford (2004): "Optimal Inflation-Targeting Rules," in *The Inflation-Targeting Debate*, National Bureau of Economic Research, Inc, NBER Chapters, 93–172.

Golub, G. H., M. Heath, and G. Wahba (1979): "Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter," *Technometrics*, 21, 215–223.

Gourieroux, C., A. Monfort, and E. Renault (1993): "Indirect Inference," *Journal of Applied Econometrics*, 8, 85–118.

Gourieroux, C., A. Monfort, and Q. Vuong (1995): *Statistics and Econometric Models*, Cambridge Books, Cambridge University Press, Cambridge, UK.

Greene, W. H. (2012): *Econometrics Analysis*, Pearson Education, London, UK.

Hall, A. R., A. Inoue, J. M. Nason, and B. Rossi (2012): "Information criteria for impulse response function matching estimation of DSGE models," *Journal of Econometrics*, 170, 499–518.

Hamilton, J. D. (1994): *Time Series Analysis*, Princeton University Press, Princeton NJ.

Hansen, C., J. Hausman, and W. Newey (2008): "Estimation With Many Instrumental Variables," *Journal of Business & Economic Statistics*, 26, 398–422.

Ireland, P. N. (2004): "A method for taking models to the data," *Journal of Economic Dynamics and Control*, 28, 1205–1226.

Kitamura, Y., G. Tripathi, and H. Ahn (2004): "Empirical Likelihood-Based Inference in Conditional Moment Restriction Models," *Econometrica*, 72, 1667–1714.

Kress, R. (1999): in *Linear Integral Equations*, Springer, vol. 82 of *Applied Mathematical Sciences*.

Magnus, J. R. and H. Neudecker (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley and Sons, Hoboken NJ.

Mandelman, F. S. and F. Zanetti (2014): "Flexible prices, labor market frictions and the response of employment to technology shocks," *Labour Economics*, 26, 94–102.

Newey, W. K. and F. Windmeijer (2009): "Generalized Method of Moments With Many Weak Moment Conditions," *Econometrica*, 77, 687–719.

Rotemberg, J. and M. Woodford (1997): "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy," in *NBER Macroeconomics Annual 1997, Volume 12*, National Bureau of Economic Research, Inc, NBER Chapters, 297–361.

Ruud, P. A. (2000): *An Introduction to Classical Econometric Theory*, Oxford University Press, Oxford, UK.

Sargan, J. D. (1983): "Identification and Lack of Identification," *Econometrica*, 51, 1605–33.

Watson, M. W. (2006): "Comment on assessing structural VARs by L. Christiano, M. Eichenbaum, and R. Vigfusson," *NBER Macroeconomics Annual*, 21, 1–105.

Zanetti, F. (2008): "Labor and investment frictions in a real business cycle model," *Journal of Economic Dynamics and Control*, 32, 3294–3314.